

R code for Data Science for Beginners

Day 4: Individual Exercise

AUTHOR

Graham Jones

PUBLISHED

September 20, 2024

Clean up your workspace

```
rm(list=ls(all=TRUE)) # remove all the named objects visible in the environment  
cat("\014") # clean your console
```

1. Let's do more exercises with dplyr (with a different dataset)

Please download the nycflights13 data by installing this package called `nycflights13`

```
#install.packages("pillar")
#install.packages("nycflights13")
library("nycflights13")
```

1-1: Please find all March flights in the data (the dataset is named "flights") flights

```
library(dplyr)
```

Attaching package: 'dplyr'

The following objects are masked from 'package:stats':

`filter`, `lag`

The following objects are masked from 'package:base':

`intersect`, `setdiff`, `setequal`, `union`

```
march_flights <- filter(flights, month == 3)
```

1-2 :Create a new variable as date with a format like this 1/1/2013, using the `mutate()` function

```
flights <- flights %>% mutate(date = paste(month, day, year, sep = "/"))
```

1-3: Change column name tailnum to tail_number

```
flights <- flights %>% rename(tail_number = tailnum)
```

1-4: Group flights by their origins

```
flights_by_origin <- flights %>% group_by(origin)
```

1-5: Count how many flights departing from JFK on 2013-12-31?

```
flights %>% filter(origin == "JFK", year == 2013, month == 12, day == 31) %>% summarise
```

```
# A tibble: 1 × 1
  n
```

```
<int>  
1    283
```

1-6: Calculate the average hours of delay in departure for all flights from JFK

```
flights %>% filter(origin == "JFK") %>% summarise(avg_dep_delay_hours = mean(dep_delay
```

```
# A tibble: 1 × 1  
  avg_dep_delay_hours  
          <dbl>  
1             0.202
```

Finally, execute the entire contents of this file. Make sure that you don't get any error message. If you get an error message, it's probably because you forgot to comment out something.